Data Collection and Preprocessing Phase

Date	10 July 2024
Team ID	SWTID1720001058
Project Title	Panic Disorder Detection
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	**Overview:** The machine learning project aims to develop a robust model for the detection of panic disorder using computational techniques and data analytics. Panic disorder is characterized by recurrent, unexpected panic attacks, which can severely impact an individual's quality of life if left untreated. **Objectives:** **Accurate Detection**: Develop machine learning algorithms capable of accurately identifying patterns and symptoms indicative of panic disorder based on input data, which may include symptom profiles, demographic information, and potentially physiological data. **Early Intervention**: Enable early detection of panic disorder to facilitate timely intervention and treatment. Early identification can help reduce the severity of symptoms and improve overall patient outcomes by providing timely access to appropriate healthcare resources.

	Personalized Diagnosis: Implement machine learning models that can adapt to individual variations in symptoms and demographics, thereby providing personalized diagnostic insights. This approach aims to enhance diagnostic accuracy compared to traditional diagnostic methods **Integration and Deployment**: Integrate the developed models into practical applications, such as clinical decision support systems or digital health platforms, to assist healthcare providers in diagnosing panic disorder more efficiently and effectively. **Ethical Considerations**: Address ethical implications related to patient privacy, consent, and fairness in algorithmic decision-making throughout the project lifecycle. Ensure that the deployment of machine learning models aligns with ethical standards and regulatory guidelines in healthcare.
Data Collection Plan	https://www.kaggle.com/datasets/muhammadshahidazeem/panic-disorder-detection-dataset
Raw Data Sources Identified	• **Clinical Records**: Structured data from electronic health records (EHRs) containing patient demographics, medical history, and diagnostic codes related to panic disorder and other mental health conditions. • **Symptom Surveys**: Self-reported or clinician-recorded surveys capturing detailed information about panic attack symptoms, frequency, duration, and associated triggers. • **Physiological Measurements**: Biometric data such as heart rate variability (HRV), skin conductance, and respiratory rate obtained through wearable devices or clinical monitoring equipment during panic attacks or controlled environments. • **Behavioral Observations**: Qualitative data from clinical observations or patient interviews detailing behavioral manifestations during panic attacks and their impact on daily functioning. • **Textual Descriptions**: Unstructured data from patient narratives, psychiatric evaluations, or therapy session transcripts

	describing subjective experiences, emotional responses, and contextual factors related to panic disorder. • **Genetic and Environmental Factors**: Research data on genetic predispositions, familial history of anxiety disorders, and environmental stressors potentially influencing the onset and severity of panic disorder. • **Public Health Datasets**: Epidemiological studies, national health surveys, or registries providing population-level statistics on prevalence rates, geographic variations, and demographic patterns associated with panic disorder.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	Panic_Disorder_training: use this file to train your model. this file contains 100000 labeled records.	https://www.kaggle.com/datasets/muhammadshahidazeem/panic-disorder-detection-dataset	CSV	2MB	Public
Dataset 2	Panic_Disorder_training: use this file to train your model. this file contains 100000 labeled records.	https://www.kaggle.com/datasets/muhammadshahidazeem/panic-disorder-detection-dataset	Excel	2MB	Private (with access)

…	…	…	…	…	…